CLAIMS

What is claimed is:

1. A method for detecting an electronic communication sent or received by a user and relating to an unsafe behavior, comprising:

analyzing one or more features of the electronic communication, said one or more features indicative of the unsafe behavior;

categorizing the electronic communication as relating to the unsafe behavior as a function of the analyzed features;

generating a report if the electronic communication is categorized as relating to the unsafe behavior, said report indicative of the unsafe behavior;.and

sending the report to a responsible person of the user.

2. The method of claim 1 further comprising specifying a type of the unsafe behavior to detect whether the electronic communication relates to said type of the unsafe behavior.

3. The method of claim 2, wherein said type of the unsafe behavior includes one or more of the following: sexual predatory behavior, bullying behavior, offensive language, personal information solicitation or revelation, personal characteristics solicitation or revelation, meeting arrangement with a stranger, picture sharing with a stranger, or a combination thereof.

4. The method of claim 3, wherein analyzing one or more features of the electronic communication further comprises:

parsing the electronic communication to generate a plurality of constituent tokens of the electronic communication;

generating, in response to said tokens, a feature vector associated with the electronic communication, said feature vector indicating whether each one of a predefined set of one or more features relating to the type of the unsafe behavior is present in the electronic communication; and

applying the generated feature vector to a probabilistic classifier relating to the type of the unsafe behavior to generate a rating for the electronic communication, wherein the probabilistic classifier trained on a content, anti-content, and close anti-content of the type of the unsafe behavior to identify said predefined set of one or more features, said rating indicating a probability that the electronic communication relates to the type of the unsafe behavior; and

wherein said categorizing the electronic communication comprises categorizing the electronic communication as relating to the type of the unsafe behavior as a function of the rating.

5. The method of claim 4, wherein the electronic communication is categorized as relating to the type of the unsafe behavior if the rating is greater than a threshold level.

6. The method of claim 4, wherein the probabilistic classifier comprises one or more classifiers selected from a group comprising: a Naïve Bayesian classifier, a limited dependence

Bayesian classifier, a Bayesian network classifier, a decision tree, a support vector machine, a content matching classifier, or a combination thereof.

7. The method of claim 4, wherein parsing the electronic communication comprises one or more of the following: parsing the electronic communication with respect to a certain period of time, parsing the electronic communication with respect to a certain size window, or parsing the entire electronic communication.

8. The method of claim 4 further comprising updating one or more of the following: the probabilistic classifier or the predefined set of one or more features.

9. The method of claim 1, wherein the report is further sent to the user.

10. The method of claim 1 further comprising not to generate and send the report if an input indicating that a source or recipient of the electronic communication is trustworthy is received from the responsible person.

11. The method of claim 1, wherein the report comprises one or more of the following: a content of the electronic communication, an identification of a source or recipient of the electronic communication, a time of the electronic communication, a type of the unsafe behavior related to the electronic communication, and a recommendation of how to address said type of the unsafe behavior.

12. The method of claim 1 further comprising indicating that the electronic communication is being analyzed to one or more of the following: a source or recipient of the electronic communication or the user.

13. The method of claim 1 further comprising generating an alert if the electronic communication is categorized as relating to the unsafe behavior, said alert informing that the electronic communication relates to the unsafe behavior to one or more of the following: the user or the responsible person.

14. The method of claim 13, wherein said alert is generated only if the electronic communication is still in progress.

15. The method of claim 1, wherein the electronic communication comprises one or more electronic messages selected from a group comprising: an email, an instant messaging session, or a chat session.

16. One or more computer readable media having computer-executable instructions for performing the method of claim 1.

17. A system adapted to detect an electronic communication sent or received by a user and relating to an undesired behavior, comprising:

a computer to receive or send the electronic communication;

computer-executable instructions to analyze one or more features of the electronic communication, said one or more features indicative of the undesired behavior;

computer-executable instructions to categorize the electronic communication as either relating to the undesired behavior or relating to an innocuous behavior as a function of the analyzed features;

computer-executable instructions to generate a report if the electronic communication is categorized as relating to the undesired behavior, said report indicative of the undesired behavior; and

computer-executable instructions to send the report to a responsible person of the user.

18. The system of claim 17 further comprising computer-executable instructions to specify a type of the undesired behavior to detect whether the electronic communication relates to said type of the undesired behavior.

19. The system of claim 18, wherein said type of the undesired behavior includes one or more of the following: sexual predatory behavior, bullying behavior, offensive language, personal information solicitation or revelation, personal characteristics solicitation or revelation, meeting arrangement with a stranger, picture sharing with a stranger, or a combination thereof.

20. The system of claim 19, wherein the computer-executable instructions to analyze one or more features of the electronic communication further comprise computer-executable instructions to:

parse the electronic communication to generate a plurality of constituent tokens of the electronic communication;

generate, in response to said tokens, a feature vector associated with the electronic communication, said feature vector indicating whether each one of a predefined set of one or more features relating to the type of the undesired behavior is present in the electronic communication; and

apply the generated feature vector to a probabilistic classifier relating to the type of the undesired behavior to generate a rating for the electronic communication, wherein the probabilistic classifier trained on a content, anti-content, and close anti-content of the type of the undesired behavior to identify said predefined set of one or more features, said rating indicating a probability that the electronic communication relates to the type of the undesired behavior; and

wherein said computer-executable instructions to categorize the electronic communication comprises computer-executable instructions to categorize the electronic communication as either relating to the type of the undesired behavior or relating to the innocuous behavior as a function of the rating.

21. The system of claim 20, wherein the electronic communication is categorized as relating to the type of the undesired behavior if the rating is greater than a threshold level.

22. The system of claim 20, wherein the probabilistic classifier comprises one or more classifiers selected from a group comprising: a Naïve Bayesian classifier, a limited dependence Bayesian classifier, a Bayesian network classifier, a decision tree, a support vector machine, a content matching classifier, or a combination thereof.

23. The system of claim 20, wherein the computer-executable instructions to parse the electronic communication comprise computer-executable instructions to perform one or more of the following: parsing the electronic communication with respect to a certain period of time, parsing the electronic communication with respect to a certain size window, or parsing the entire electronic communication.

24. The system of claim 20 further comprising computer-executable instructions to update one or more of the following: the probabilistic classifier or the predefined set of one or more features.

25. The system of claim 17, wherein the report is further sent to the user.

26. The system of claim 17 further comprising computer-executable instructions not to generate and send the report if an input indicating that a source or recipient of the electronic communication is trustworthy is received from the responsible person.

27. The system of claim 17, wherein the report comprises one or more of the following: a content of the electronic communication, an identification of a source or recipient of the electronic communication, a time of the electronic communication, a type of the undesired behavior related to the electronic communication, and a recommendation of how to address said type of the undesired behavior.

28. The system of claim 17 further comprising computer-executable instructions to indicate that the electronic communication is being analyzed to one or more of the following: a source or recipient of the electronic communication or the user.

29. The system of claim 17 further comprising computer-executable instructions to generate an alert if the electronic communication is categorized as relating to the undesired behavior, said alert adapted to inform that the electronic communication relates to the undesired behavior to one or more of the following: the user or the responsible person.

30. The system of claim 29, wherein said alert is generated only if the electronic communication is still in progress.

31. The system of claim 17, wherein the electronic communication comprises one or more electronic messages selected from a group comprising: an email, an instant messaging session, or a chat session.

32. The system of claim 17, wherein said computer is a server or a client.

33. The system of claim 17, wherein said computer is a client and said computer-executable instructions to analyze the electronic communication are located on a server, and wherein the server is adapted to receive the electronic communication sent or received by the client for said computer-executable instructions located on the server to analyze the electronic communication.

34. A computer-readable medium having computer-executable instructions for performing a method to detect an electronic communication sent or received by a user and relating to an unsafe behavior, said method comprising:

analyzing one or more features of the electronic communication, said one or more features indicative of the unsafe behavior;

categorizing the electronic communication as relating to the unsafe behavior as a function of the analyzed features;

generating a report if the electronic communication is categorized as relating to the unsafe behavior, said report indicative of the unsafe behavior; and

sending the report to a responsible person of the user.

35. The computer-readable medium of claim 34, wherein the method further comprises specifying a type of the unsafe behavior to detect whether the electronic communication relates to said type of the unsafe behavior.

36. The computer-readable medium of claim 35, wherein said analyzing one or more features of the electronic communication comprises:

parsing the electronic communication to generate a plurality of constituent tokens of the electronic communication;

generating, in response to said tokens, a feature vector associated with the electronic communication, said feature vector indicating whether each one of a predefined set of one or more features relating to the type of the unsafe behavior is present in the electronic communication; and

applying the generated feature vector to a probabilistic classifier relating to the type of the unsafe behavior to generate a rating for the electronic communication, wherein the probabilistic classifier trained on a content, anti-content, and close anti-content of the type of the unsafe behavior to identify said predefined set of one or more features, said rating indicating a probability that the electronic communication relates to the type of the unsafe behavior; and

wherein said categorizing the electronic communication comprises categorizing the electronic communication as relating to the type of the unsafe behavior as a function of the rating.

37. The computer-readable medium of claim 36, wherein parsing the electronic communication comprises one or more of the following: parsing the electronic communication with respect to a certain period of time, parsing the electronic communication with respect to a certain size window, or parsing the entire electronic communication.

38. The computer-readable medium of claim 34, wherein the method further comprises not to generate and send the report if an input indicating that a source or recipient of the electronic communication is trustworthy is received from the responsible person.

39. The computer-readable medium of claim 34, wherein the method further comprises indicating that the electronic communication is being analyzed to one or more of the following: a source or recipient of the electronic communication or the user.

40. The computer-readable medium of claim 34, wherein the method further comprises generating an alert if the electronic communication is categorized as relating to the unsafe behavior, said alert informing that the electronic communication relates to the unsafe behavior to one or more of the following: the user or the responsible person.